Master in Actuarial and Financial Sciences

MASTER THESIS

Title: Bitcoin: A long-run equilibrium with East Asian currencies

Author: David Anaya Luque

Advisor: Samer Ajour El Zein

Academic year: 2019-2020

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

Bitcoin: A long-run equilibrium with East Asian currencies

Author: David Anaya Luque

Advisor: Samer Ajour El Zein

"The content of this document is the sole responsibility of the author, who declares that he has not incurred plagiarism and that all references to other authors have been expressed in the text."

Abstract

The purpose of the study will be focused on developing and demonstrating the existence of cointegration between the most important cryptocurrency in the world, BTC-Bitcoin, paired with three different local currencies from selected southeast Asian countries: China (BTC-CNY), South Korea (BTC-KRW) and Japan (BTC-JPY).

In order to achieve the proposed objective, it is extensively developed the entire process currently established by the most important econometric literature, from the analysis of stationarity of the series and the detection of unit roots to the validation of a VAR model to be contrasted with the Johansen test, its resultant cointegration equation (long-run model) and the dynamic adjustment for each of the target variables.

Finally, it has been proposed an ARDL bounds testing approach model to determine and contrast the results obtained with the Johansen test and ensure the existence of a real long-term relationship of the selected series, where the coefficients obtained that determine the link between integrated processes and steady state equilibrium (no growth steady state) using Johansen approach are: 0.0056$(lnBTCKRW_t)$ and 0.0037$(lnBTCJPY_t)$; and through the use of an ARDL bounds testing approach the following coefficients are obtained: 0.00556$(lnBTCKRW_t)$ and 0.00426$(lnBTCJPY_t)$.

Keywords: cryptocurrency, bitcoin, econometric, financial , cointegration,

Contents

Abstract **i**

1 Introduction **1**
1.1 Thesis goals 2

2 Theoretical and conceptual framework **3**

3 Econometric methodology, modeling and estimation **7**
3.1 Data 8
3.2 Financial time series 8
3.2.1 Time series analysis 8
3.2.2 Econometric model building 15
VAR(p) model 15
Granger causality 18
Impulse - Response function 20
3.2.3 Cointegration tests 22
Johansen test 24
ARDL Bounds Testing approach 26

4 Conclusions and research limitations **30**
4.1 Conclusions 30
4.2 Limitations of the study 32

A Additional ADF, PP and KPSS tables **34**

B Orthogonal Impulse Response **36**

C Short-run response coefficients **39**

D Plotting VECM model responses **42**

E ARDL model coefficients **43**

F RECM coefficient values **44**

G ARDL - Cointegration equation **45**

Bibliography **46**

List of Figures

2.1 Total Market Capitalization 5
2.2 Percentage Total Market Capitalization 6

3.1 Bitcoin price series/returns against local currency 8
3.2 ACF - PACF for regular Bitcoin price series 11
3.3 ACF - PACF for Bitcoin return series 11

B.1 Orthogonal Impulse Response from BTCCNY 37
B.2 Orthogonal Impulse Response from BTCKRW 37
B.3 Orthogonal Impulse Response from BTCJPY 38

C.1 Response BTCCNY coefficients 40
C.2 Response BTCKRW coefficients 40
C.3 Response BTCJPY coefficients 41

D.1 VECM model responses 42

E.1 ARDL model coefficients 43

F.1 Short-run dynamic and long-run coefficients 44

G.1 Cointegration equation - Projected vs Observed 45

List of Tables

2.1 Percentage of Total Market Capitalization 4

3.1 ADF, PP and KPSS tests of BTC-CNY series (prices) 13
3.2 ADF, PP and KPSS tests of BTC-CNY series (returns) 13
3.3 Zivot-Andrews test of BTC-CNY series (prices) 14
3.4 Zivot-Andrews test of BTC-CNY series (returns) 14
3.5 VAR(p) model lag selection 17
3.6 Multivariate Portmanteau Test (adjusted) 18
3.7 Granger causality VAR(1), m=3 19
3.8 Variance - covariance matrix VAR(1), m=3 20
3.9 Correlation matrix VAR(1), m=3 20
3.10 OR - Choleski decomposition VAR(1), m=3 21
3.11 Trace test - VAR(p=1), k=2 24
3.12 Maximum eigenvalue test - VAR(p=1), k=2 25
3.13 Bounds F-test (Wald) 28
3.14 ARDL model selection 28
3.15 Durbin-Watson test 28
3.16 Breusch-Pagan test 29
3.17 Ramsey RESET test 29

A.1 ADF, PP and KPSS tests of BTC-KRW series (prices) 35
A.2 ADF, PP and KPSS tests of BTC-KRW series (returns) 35
A.3 ADF, PP and KPSS tests of BTC-JPY series (prices) 35
A.4 ADF, PP and KPSS tests of BTC-JPY series (prices) 35

Chapter 1

Introduction

Cryptocurrencies (Lam and Lee, 2015) are a subset of the class of digital currencies that have gone from being a digital payment instrument to a great source of interest when it comes to making large-volume investments and speculating on their evolution in the markets. Taking advantage of the wide range of econometric methods that are currently used to manage financial asset portfolios, the aim of this study is to focus on verify the existence of cointegration of the selected three series of Bitcoin in order to determine the long-run model and the short-run dynamics adjustments that explain the model.

In order to detail the progress that it has achieved and what is being done at every stage, it has been decided to organize the project as follows:

- In **Chapter 2** it is exposed a brief approximation of the theoretical and conceptual framework that exists around Bitcoin and general cryptocurrencies in the market, the volume of capitalization and the level of dominance of each of them in the actual market.

- Once they are introduced, **Chapter 3** introduces the econometric methodology, modeling and estimation that is going to be established in order to achieve the objective of this work, where there are exposed the main tests and contrasts applied during the econometric process and the methodological development behind each one.

- Finally, **Chapter 4** introduces the final conclusions and research limitations observed during the econometric process of determining our main objectives, identifying literature gaps and the main problems that could currently alter the results of our study.

1.1 Thesis goals

The goals of the thesis are dual. The thesis is organized to commence with an extensive literature review of the time series analysis that comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. In addition, it also focuses on that literature that involve the whole process of cointegration analysis, the evidence that exists and alternative methods to study this relationship in the long-run state. Everything followed by an explanatory description of the project objectives and methodology.

To meet the second goal, all the literature analyzed is applied to schematically develop the whole econometric process that includes the determination of cointegration equations and the dynamic adjustments using the Johansen test and, comparatively, deriving and modelling an ARDL model using the bounds testing approach. This process, together with the selected literature review, allows us to derive robust results about the hypotheses developed in this study and also to confirm that the goals of the thesis have been met opening the line for a further research.

Chapter 2

Theoretical and conceptual framework

Nowadays it is hard to imagine someone who has never heard before the word Bitcoin or cryptocurrency in any media such as television or economic newspapers. That is because in 2009 someone decided to create an open-source software to break with the traditional banking system and let a digital asset flow away through the internet. That open-source was called Bitcoin and it was considered the first decentralized cryptocurrency and distributed ledger technology, typically well-known as a blockchain. It couldn't be copied, it couldn't be duplicated or pirated, so those primary attributes gave its confidence and secure to the people that wanted to get in this world.

Since then, many cryptocurrencies have been created as a medium of exchange to do secure financial transactions in internet outside the existing banking system. Even to use them as a stock exchange in the market, most of them paired with the most important currencies in the world and they have generated unimaginable profits to all those people who decided to go long on these currencies in its first years of existence.

By definition, and according to (Narayanan and Goldfeder, 2016), digital currencies are, unlike physical currencies, not controlled by economic systems and centralized banks. It is also an anonymous ecosystem, which means that no one can identify who is making transactions while still making the transaction data public. This makes it protected from third parties and lobbies that could influence the demand or control the prices, and that is why certain politicians from some different countries have banned the use of this type of digital currencies over the said territory. The rate with which new currency is created is defined from the beginning and is known to the public.

At the same time, though, most of them have also been designed to progressively decrease the production of new units -making it reach a cap on the total amount of currency that exists in order to prevent high inflation.

At present, according to *CoinMarketCap*[1] reports, there are more than 5.000 **cryptocurrencies** in the world and more than 21.000 **markets**, where these digital currencies are paired with standard reference currencies and in most of them the US dollar predominates. **Total market capitalization** average stands around $ 250.000.000.000, and the **market share** or dominance of the most important cryptocurrencies over the total market capitalization is shown in the following table, where it can be clearly seen how **Bitcoin** accounts for almost 64% of the total.

TABLE 2.1: Percentage of Total Market Capitalization, "Source: own elaboration"

Cryptocurrency	% of total
Bitcoin	63.77 %
Ethereum	9.87 %
XRP	3.91 %
Tether	2.87 %
Bitcoin Cash	2.04 %
Bitcoin SV	1.62 %
Litecoin	1.30 %

The following graphic (Figure 1.1) shows the time series as a whole with respect to the total market capitalization of each of the cryptocurrencies, where the volume remains constant during the previous years to 2016 due to the low opening of this type of currency to the world and the technological environment.

Additionally, it must be taken into account that, for example, Bitcoin was born in 2009 and was one of the youngest to appear in this environment, but most of the 5,000 that currently exist have been born in mid-2013, which makes them have a fairly tight margin to grow in a market where Bitcoin has been made with practically all of it.

[1]https://coinmarketcap.com/all/views/all/

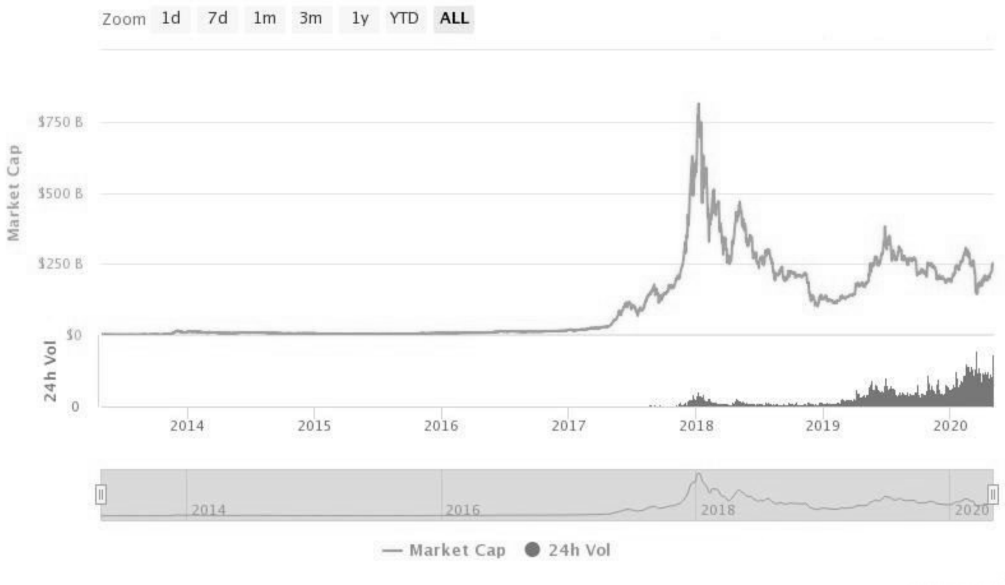

FIGURE 2.1: Total Market Capitalization,
"Source: CoinMarketCap"

Even so, the significant growth that Bitcoin showed did what it could be said today as the starting point of everything that exists right now, and what is to come.

New digital currencies began to emerge, new blockchain-related technologies that further improved the anonymity and security of participants in the network, new protocols that were created allowed more efficient mining, in environments such as Tor, and therefore, more profit and more speculation.

Two of the cryptocurrencies that managed to move important *whales*[2] and private investment funds to break the uptrend of Bitcoin in 2017 were **Ethereum** and **XRP**.

Both were born in 2013 and 2012 respectively and thanks to the development of new technologies, they managed to be one of the currencies with the highest gains in 2017 [3], and large-scale displacement of an important market volume towards them.

[2]https://btcmanager.com/what-are-whales-how-do-they-affect-cryptocurrencies/
[3]https://www.coindesk.com/900-20000-bitcoins-historic-2017-price-run-revisited

Because of that, as it is explained in (Fry and Cheah, 2016), the prices of the currencies in the markets began to move very quickly, causing great profits and at the same time great losses to large investors that didn't know why volatility had shot up so aggressively, as that large shifting of currencies from portfolio to portfolio was displacing supply and demand out of control. Fortunately, time has made it possible for these situations to be foreseen as far as possible, and there are web pages that centralize the registration of all transactions in real time (which are public as explained above) to detect ***anomalies***[4] or important movements of currencies between wallets that can shake the whole system. Although the origin or final destination of a transaction can be unknown, with the data of each movement available to anyone who wants information about how much money is flowing through the network at a certain point in time, the system may be somewhat more transparent against the excess of anonymity.

The next graphic shows how Bitcoin's market dominance has been declining since 2013 and, especially, the great debacle in late 2017 and early 2018 in a very graphic way.

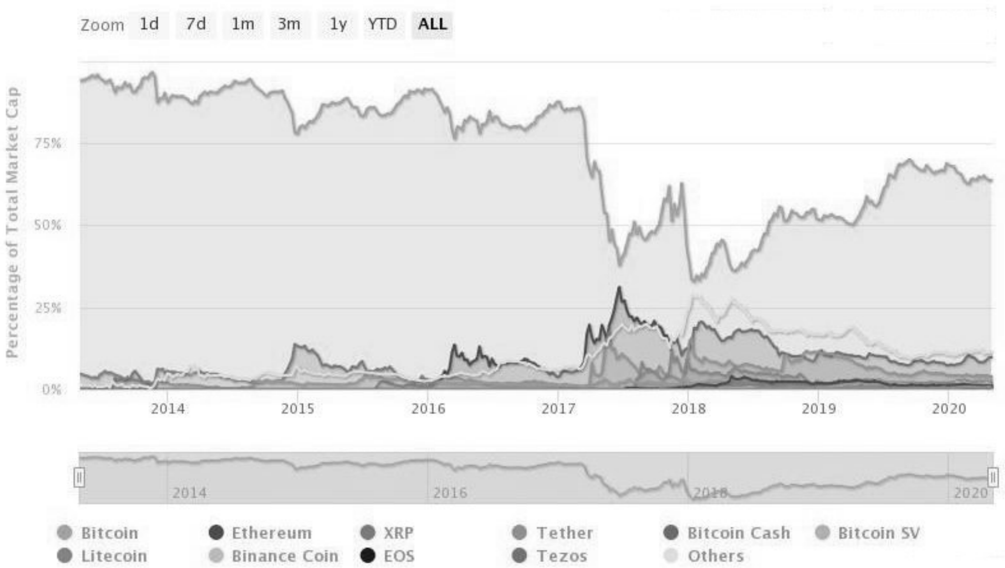

FIGURE 2.2: % of Total Market Capitalization,
"Source: CoinMarketCap"

[4] https://whale-alert.io/

Chapter 3

Econometric methodology, modeling and estimation

One of the aspects that is being very important in this field especially to explain certain pairs of economic or financial variables is what is known as the existence of **cointegration**, formalized by (Engle and Granjer, 1987), and approached as a long-run economic relationship explained by a linear combination I(0) of two or more variables I(1). In the following sections to be developed - within the econometric environment with the aim of detecting the movement pattern of each of the series analyzed and its behavior in the long term, as well as the short-term shocks that may alter the steady state, a study based on demonstrating the existence or not of cointegrated series and the entire scaled process to develop the study will be considered, including both theoretical and practical approaches.

First steps are focused on analyzing the time series data - a summary of the main descriptive statistics for each series, verifying the adequacy of them using techniques to ensure the existence of stationarity and non correlated residuals. Following to the next steps, it is going to be focused on determining an optimal VAR(p) model through the use of different information criterion and verifying the adequacy of the model but this time using multivariate tests. In addition, the Impulse-Response function and the Granger causality test are applied to describe the evolution of the model's variables in reaction to a shock or to determine whether one time series is useful in forecasting another.

Finally, it is covered the entire process derived from determining the calculations related to the resultant cointegration equation. In this section, it is first exposed the Johansen method, and later on, the determination of an ARDL Bounds testing approach model as an alternative method to compare the obtained results and verify that the results obtained are consistent and may or may not justify the existence of co-integrated series.

3.1 Data

To begin the analysis, the financial time series referring to the price of Bitcoin will be used at par with the local currency of each country. So in total there will be three series with the following labels: **BTC-CNY** for the Chinese case; **BTC-KRW** in the case of South-Korean; and **BTC-JPY** for the Japanese case. The sequence of discrete-time data will take place from (2014.9) to (2019.9), with a daily observation frequency (base 365.25). The data has been extracted from the Yahoo! Finance using the Quantmod package in R.

The reason why Bitcoin is paired with the local currency is because it is desired to observe if there any macroeconomic implication or impact from the current national currency from each of the countries in our estimations.

3.2 Financial time series

3.2.1 Time series analysis

The first step will consist of analyzing the data available for each of the series and performing the first diagnostic tests to determine if they are series that will be adapted to our objectives. To do this, it has to be identified if it is necessary to stabilize mean and variance using logarithmic differentiation.

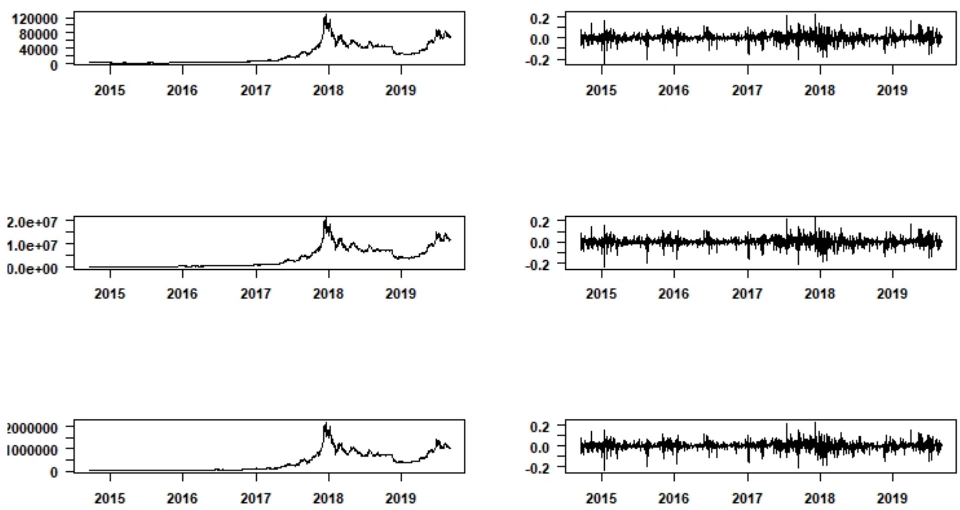

FIGURE 3.1: Bitcoin price series/returns against local currency, "Source: own elaboration"

The three plots on the left display the prices of each of the series and how invariable they remain from late-2014 until mid-2016. From early 2017 they rise exponentially from a global average of 1-3 dollars to an average of 400-1500 dollars, with average growth rates of 20,000% and 30,000%, respectively.

Once they top the maximum resistance in mid-2018, they reverse immediately to the initial average prices as a consequence of the massive closing long positions.

Then, it can be affirmed that they are not stationary, because they present a variability that increases with the level of the series and their behaviour keep changing over time in an unpredictable way. This is what is known as heteroskedasticity.

The other three plots on the right display the returns, where it can be anticipated that these, just as it should be, will be stationary and integrated of order 0, because it is known that all stationary processes are integrated **I(0)**, but not all **I(0)** processes are stationary.

It is clear the existence of relationship between each of the series: same pattern in the evolution of the price and range of similar returns, with no clearly defined volatility clusters over the analyzed period of time. The following tables contain the descriptive statistics for each series.
Before starting to execute commands, it is necessary to make an analysis of the most important statistics of our data to ensure that there are no errors, and in the same way, to better know what type of observations are selected.

The following tables contain the descriptive statistics for each series:

basicStats BTC-CNY

NAs	0
Mean	520660.2
Stdev	16793.51
Skewness	-0.096634
Kurtosis	-0.496225

basicStats BTC-KRW

NAs	0
Mean	2966433
Stdev	4086366
Skewness	1.739114
Kurtosis	2.554908

basicStats BTC-JPY

NAs	0
Mean	298831
Stdev	418792.3
Skewness	1.824194
Kurtosis	3.000417

If the tables are analyzed, it can be seen that there is no NA values, therefore, it can be ensured that we do not have incomplete data in our series. In the same way, it can be seen that the skewness is not 0, it is greater(lower) than this value, but with levels very close to 0, therefore, it will differ in terms of data size in both queues, but a very small difference. It can be applied the same criteria in the case of kurtosis.

To continue with the process of verifying the stationarity of the time series data, it will be proceeded to start **informal methods** to evaluate the existence of unit root.

To do that, it is going to calculate the **ACF** and **PACF**, which is known as Autocorrelation Function and Partial Autocorrelation Function that it is introduced by (Dickey and Fuller, 1979), in order to detect the linear dependence of a variable with itself at two points in time.

By definition, it is known that for stationary processes, autoccorrelation between any two observations only depends on the time lag **h** between them. Additionally, for testing the significance of a single lag-h autocorrelation it is calculated the standard error.

So, mathematically:

$$\rho_h = Corr(y_t, y_t - h) = \frac{\gamma_h}{\gamma_o} \tag{3.1}$$

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^{T}(y_t - \hat{y})(y_{t-h} - \hat{y})}{\sum_{t=1}^{T}(y_t - \hat{y})^2} \tag{3.2}$$

$$SE_p = \sqrt{(1 + \frac{2\sum_{i=1}^{h-1}\hat{\rho}_i^2}{N}} \tag{3.3}$$

On the next page there are exposed the outputs obtained by applying this procedure:

- Figure 2.2 shows how the price series do not look stationary in each of the ACF plots and it can be seen how they remain constant and fall very slowly over the lags period.

- At the same time, Figure 2.3 shows in the opposite way how the series of returns fall sharply in the ACF and PACF plots, meaning that only describe the direct relationship between an observation and its lag.

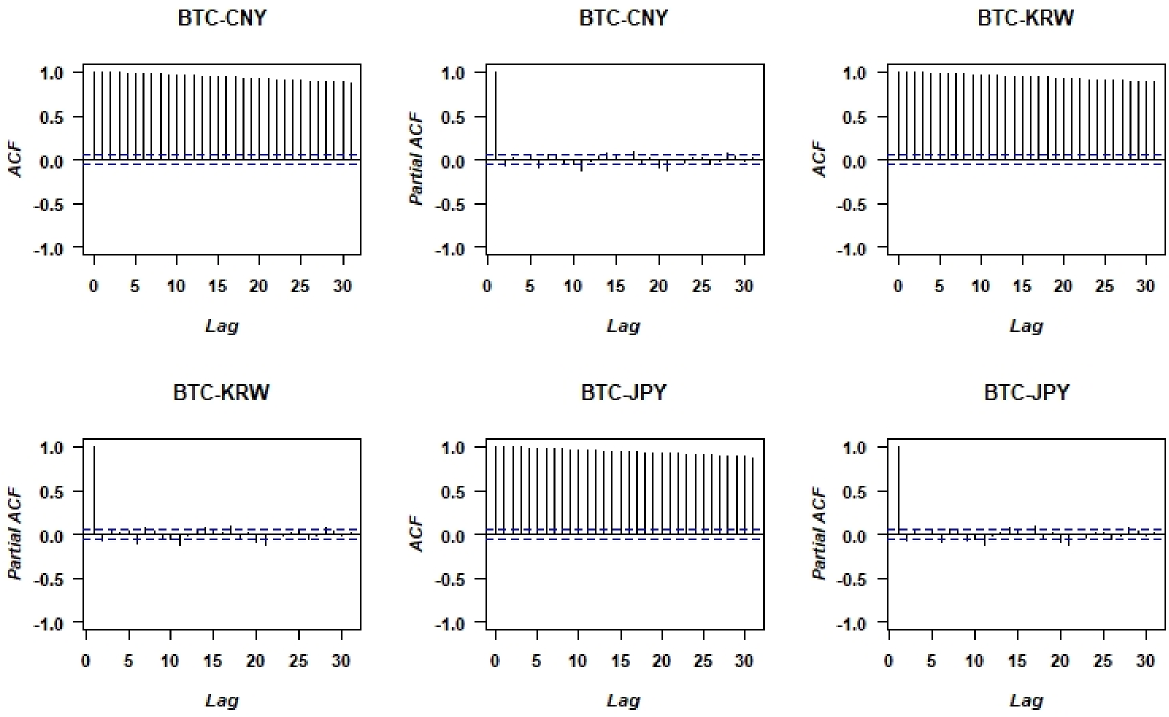

FIGURE 3.2: ACF - PACF for regular Bitcoin price series

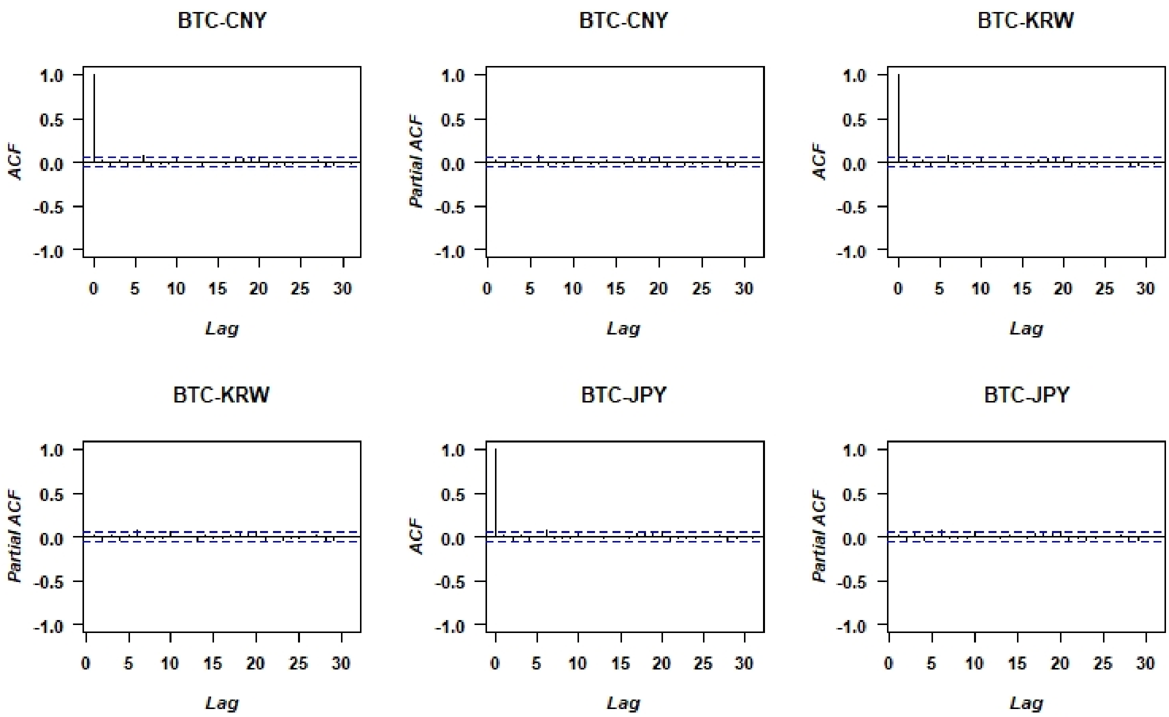

FIGURE 3.3: ACF - PACF for Bitcoin return series

To make easier the interpretation, (Ljung and Box, 1978) introduced an aproximation of what is known as **Ljung-Box Q-test**, that it is also applied to check, not just visually, the null hypothesis that a series of residuals exhibits no autocorrelation for a fixed number of lags, against the alternative hypothesis that some autocorrelation coefficient is different from zero - they should be near zero for any and all time-lag.

$$Q = T(T+2)\sum_{K=1}^{L}(\frac{\rho(k)^2}{(T-k)} \tag{3.4}$$

As expected, once it is computed, it is clear that it is necessary to transform the time series data in order to meet the first requirement, that is stationarity. To achieve this, logarithms will be applied to correct the heteroscedasticity, and it will be differentiated to eliminate the trend and the seasonal component.

But, before carrying out these transformations, there are other ways to prove that the series really have to be differentiated, and they are what are called as **formal methods**.

- **Augmented Dickey-Fuller test (ADF)**: to test the null hypothesis that exist a unit root under the approximation of the critical value by (Dickey and Fuller, 1979).
- **Phillips-Perron test (PP)**: to test the null hypothesis that a time series is I(1), introduced later than the previous methodology and developed by (Phillips and Perron, 1988) in their publication.
- **Kwiatkowski-Phillips-Schmidt-Shin test (KPSS)**: to test the same but as H1. This one is the last tests introduced by (Kwiatkowski and Shin, 1992) with some additions that differs from the previous ones.

The first of the tests, the **ADF**, tests the null hypothesis from the following expression:

$$y_t = c + \delta_t + \emptyset y_{t-1} + \beta_p \partial y_{t-p} + \epsilon_t, \tag{3.5}$$

where the null hypothesis is empty set is equal to 1 (exist a unit root), and the alternative hypothesis is that empty set is strictly less than 1.
The second of the tests, the **PP**, tests the null hypothesis from the following expression:

$$y_t = c + \delta_t + \emptyset y_{t-1} + \epsilon(t), \tag{3.6}$$

where the alternative hypothesis is that a time series is not integrated of order

1, and the null hypothesis is that the time series is integrated of order 1, I(1), under the equation:

$$(1 - L)^d X_t, \ d = 1 \tag{3.7}$$

Finally, the third of the tests, the **KPSS**, test the inverse null hypothesis of the PP test following expression:

$$y_t = c_t + \delta_t + u_{1t} \tag{3.8}$$

$$c_t = c_{t-1} + u_{2t} \tag{3.9}$$

The following tables collect the results of applying each of the tests in each of the series, with the corresponding p-value (t-value) and the t-statistic value.

TABLE 3.1: ADF, PP and KPSS tests of BTC-CNY series (prices), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	0.1699	-1.2246	2.1867
T-statistic value	-	-2.863999	0.146
Results	NRH0	NRH0	RH0

TABLE 3.2: ADF, PP and KPSS tests of BTC-CNY series (returns), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	0.01	-36.8573	0.0056
T-statistic value	-	-2.864016	0.146
Results	RH0	RH0	NRH0

It is only shown for **BTC-CNY** time series, but the pattern is the same for all series. Once the series are differentiated, they are stationary, and if it is plotted all three differentiated series in just one plot it can be observed that they are practically identical. Even so, the tables with the results for the remaining two variables are attached in **Appendix B**, for both price series and returns.

Additionally, it can be applied the **Zivot-Andrews test** to contrast the null hypothesis of unit root with structural break in the intercept. Although this test is not usually included in the group of formal methods to identify the existence of unit root, it is usually used especially in economic and financial series to detect potential break points at a certain point in time.

This one developed by (Zivot and Andrews, 1992) in their publication related with the Great Crash and the Oil-Price Shock.

TABLE 3.3: Zivot-Andrews test of BTC-CNY series (prices), "Source: own elaboration"

Test	Zivot-Andrews
Null hypothesis	Non stationary
P-value/T-value	-3.2656
T-statistic value	-4.8
Results	NRH0

TABLE 3.4: Zivot-Andrews test of BTC-CNY series (returns), "Source: own elaboration"

Test	Zivot-Andrews
Null hypothesis	Non stationary
P-value/T-value	-31.1698
T-statistic value	-4.8
Results	RH0

As it is rejected the null hypothesis of unit root in favour of a one time break in the intercept, then it is assumed the alternative hypothesis of a trend stationary process with a break in the intercept at position 1186 ["2017-09-15"].

The same process can be applied for the other series and it will be achieved the same result, which return series are stationary and the same potential break point at the same position.

3.2.2 Econometric model building

VAR(p) model

The main objective of carrying out this analysis through the use of multivariate models is to determine if in the long term these three series, BTC-CNY, BTC-KRW and BTC-JPY, will tend to move in the same way over the time and if they will converge each other at the same price level. This statistical property is known as cointegration, and reflects the presence of a long-run equilibrium towards which the economic system converges over time in order to maximize forecasting methods. By definition, if there exists a stationary linear combination of the three non stationary series, the series combined are said to be cointegrated.

Once it is assured that the price series are integrated of order 1 and returns integrated of order 0, the next step would be focused on determinate a **Vector Autoregressive Model**, which is a multidimensional extension well detailed in (Hamilton, 1994) of the classical autoregressive models.
It is a model that contains a system of n equations of n distinct, stationary response variables as linear functions of lagged responses. It is characterized by their degree p, **VAR(p)**, that represents the number of lags of all variables in the system.

A general Vector Autoregressive Model is similar to the AR(p) model except that each quantity is vector valued and matrices are used as the coefficients. Taking into consider this idea, the general form of a VAR(p) model is:

$$X_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_p X_{t-p} \epsilon_t \tag{3.10}$$

where every X multiplying the beta parameter is called the **p-th lag of Xt**; α parameter is a k-vector of constants (intercepts); β parameter is a time-invariant (k x k)-matrix, and the last component is a k-vector of **error terms**.

For example: **VAR(1)** with three variables/series (**m=3**)

$$X_{1,t} = \alpha_1 + \beta_{1,1} X_{1,t-1} + \beta_{1,2} X_{2,t-1} + \beta_{1,3} X_{3,t-1} \epsilon_{1t}$$

$$X_{2,t} = \alpha_1 + \beta_{2,1} X_{1,t-1} + \beta_{2,2} X_{2,t-1} + \beta_{2,3} X_{3,t-1} \epsilon_{2t}$$

$$X_{3,t} = \alpha_1 + \beta_{3,1} X_{1,t-1} + \beta_{3,2} X_{2,t-1} + \beta_{3,3} X_{3,t-1} \epsilon_{3t}$$

For example: **VAR(1)** with (**m=3**) can be written in matrix form as:

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{bmatrix} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{bmatrix} x \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix} \tag{3.11}$$

Then, in order to select the optimal number of lags to use to estimate the model, it is going to be estimated a VAR(p) by OLS per equation using the function **VARselect()** in R programming language. This reports the appropiate number of lags using an information criteria, such as:

- **Akaike's Information Criterion**, (Akaike, 1974) :

$$AIC(n) = -2max[log(L)] + 2K, \tag{3.12}$$

where **max[log(L)]** is the maximized log likelihood function and **K** the number of parameters.

- **Hannan-Quinn Information Criterion**, (Quinn, 1980) :

$$HQ(n) = Klog(n) - 2max[log(L)], \tag{3.13}$$

where **max[log(L)]** is the maximized log likelihood function, **K** the number of parameters and **n** the number of observations.

- **Schwarz Information Criterion**, (Schwarz, 1978) :

$$SC(n) = -2max[L] + 2K * log(log(n)), \tag{3.14}$$

where **max[L]** is the maximized likelihood function, **K** the number of parameters and **n** the number of observations.

Computing the previous function taking into account the structure and its components:
The previous table details, by minimizing the result obtained by each information criterion, that the four possible alternatives suggest that the **VAR(p=1)** model is the most appropriate. (The following figures refer to the differentiated log variables)

$$BTCCNY_t = \alpha_1 + \beta_{1,1}BTCCNY_{t-1} + \beta_{1,2}BTCKRW_{t-1} + \beta_{1,3}BTCJPY_{t-1}\epsilon_{1t}$$

TABLE 3.5: VAR(p) model lag selection,
"Source: own elaboration"

VAR(p)	AIC(n)	HQ(n)	SC(n)	FPE(n)
1	-27.82954	-27.81309	-27.78548	8.199405e-13
2	-27.82031	-27.79153	-27.74321	8.275439e-13
3	-27.81302	-27.77190	-27.70287	8.335987e-13
4	-27.80608	-27.75262	-27.66289	8.394053e-13
5	-27.79716	-27.73136	-27.62092	8.469318e-13
...	...	...	...	...
10	-27.78380	-27.65631	-27.44233	8.583392e-13
Selection	1	1	1	1

$$BTCKRW_t = \alpha_2 + \beta_{2,1} BTCCNY_{t-1} + \beta_{2,2} BTCKRW_{t-1} + \beta_{2,3} BTCJPY_{t-1} \epsilon_{2t}$$
$$BTCJPY_t = \alpha_3 + \beta_{3,1} BTCCNY_{t-1} + \beta_{3,2} BTCKRW_{t-1} + \beta_{3,3} BTCJPY_{t-1} \epsilon_{3t}$$

And the coefficients for the estimated model are, in the previous order:

$$BT\widehat{CC}NY = (0.0020) + (0.0094) BTCCNY_{t-1} + (0.4200) BTCKRW_{t-1} - (0.4090) BTCJPY_{t-1}$$

$$BT\widehat{CK}RW = (0.0020) + (0.062) BTCCNY_{t-1} + (0.3328) BTCKRW_{t-1} - (0.3772) BTCJPY_{t-1}$$

$$BT\widehat{CJ}PY = (0.0019) + (0.0698) BTCCNY_{t-1} + (0.3672) BTCKRW_{t-1} - (0.4167) BTCJPY_{t-1}$$

High values of p (overestimation or over-parametrization) will reduce the forecast precision of the corresponding estimated VAR(p) model, the estimation precision of the impulse response, and the approximated mean square error matrix with each additional lag level.

For that reason, it is decided to accept p = 1 as the ideal number of lags that minimize the value of the each of the information criteria method to carry out the study with a VAR at this level/order.

But before consolidating this lag order as the appropriate value for the model, it has to be diagnosed in order to detect undesirable situations like the existence of autocorrelation of the residuals.

To do that, (Edgerton and Shukur, 1999) expose clearly what is and how to deal with the **Portmanteau Autocorrelation Test**, that computes the multivariate Box-Pierce/Ljung-Box Q-statistics for residual serial correlation up to the specified order. The diagnosis tests used in VAR models are similar to those used in AR models, but amplified to a dimension of k correlation matrix of the residuals.

To check if the VAR residual are white noise, the hypothesis to test is:

$$H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0 \qquad (3.15)$$
$$H_1 : \rho_i \neq 0$$

where the multivariate Portmanteau adjusted approximation has the following form:

$$Q_k(m) = T^2 \sum_{k=1}^{m} (\frac{1}{(T-k)} tr(\hat{\rho}_k' \hat{\rho}_0^{-1} * \hat{\rho}_k \hat{\rho}_0^{-1}) \qquad (3.16)$$
$$\hat{\rho}_k = \hat{\rho}_{ij}(k)$$

Computationally, the null hypothesis of no residual autocorrelation up to lag m is tested to check how many of them pass the test. In this case, as the series are daily observed over the analyzed period, it is contrasted the VAR(p=1) until the lag 14 (2 weeks seems enough) to test for serially correlated errors.

$$H_0 : \rho_2 = \rho_3 = \rho_4 = \rho_5 = \rho_6 = ... = \rho_{14} = 0 \qquad (3.17)$$

The following table shows that as the p-value is in the acceptance region, it cannot be rejected with a significance of 5% the null hypothesis of no residual autocorrelation of VAR(1) up to lag 14.

TABLE 3.6: Multivariate Portmanteau Test (adjusted), "Source: own elaboration"

Portmanteau Test (adjusted)	**Residuals of VAR**
Chi-squared	117.72
df	117
p-value	0.4639
Result	NRH0

Granger causality

The (Granger, 1969) **causality test** is a statistical hypothesis test for determining whether one time series is useful in forecasting another. That means that, in Granger's sense, it shows if the past values of one of the time series analyzed Granger cause the other two series; or in other words, if the first time series helps to forecast one of the other two series.

The idea of this test is that there may exist co-movements between the three time series (variables) - they will trend together in finding a cointegrating relationship, then there is causality between these variables at least in one direction. It can exists unidirectional causality relationships, a bidirectional causality relationships between them, or lack of any causal relationships.

For example, assuming the estimated **VAR(1)**, the null hypothesis is such as:

TABLE 3.7: Granger causality VAR(1), m=3,
"Source: own elaboration"

Granger causality	Coefficient	p-value
BTCCNYt does $\not\xrightarrow{Granger}$ BTCKRWt	$\beta_{2,1} = 0$	0.9486
BTCKRWt does $\not\xrightarrow{Granger}$ BTCCNYt	$\beta_{1,2} = 0$	0.06315
BTCJPYt does $\not\xrightarrow{Granger}$ BTCKRWt	$\beta_{2,3} = 0$	0.0525
BTCKRWt does $\not\xrightarrow{Granger}$ BTCJPYt	$\beta_{3,2} = 0$	0.06315
BTCJPYt does $\not\xrightarrow{Granger}$ BTCCNYt	$\beta_{1,3} = 0$	0.0525
BTCCNYt does $\not\xrightarrow{Granger}$ BTCJPYt	$\beta_{3,1} = 0$	0.9486

Using a **robust heteroscedasticity-consistent** variance-covariance matrix for the Granger test, in every resultant p-value from the previous table of each relationships it can be said with a significance of 5% that:

$$\text{BTCCNYt does} \not\xrightarrow{Granger} \text{BTCKRWt}$$
$$\text{BTCKRWt does} \not\xrightarrow{Granger} \text{BTCCNYt}$$
$$\text{BTCJPYt does} \not\xrightarrow{Granger} \text{BTCKRWt}$$
$$\text{BTCKRWt does} \not\xrightarrow{Granger} \text{BTCJPYt}$$
$$\text{BTCJPYt does} \not\xrightarrow{Granger} \text{BTCCNYt}$$
$$\text{BTCCNYt does} \not\xrightarrow{Granger} \text{BTCJPYt}$$

Therefore, it can be affirmed that none of the paired variables can affect the other variables of the same vector and does not provide anything significant when it comes to predicting at some stage in the future.

Impulse - Response function

Since all variables in a VAR model depend on each other, individual coefficient estimates only provide limited information on the reaction of the system to a shock. Then, here appears the **IR function**, popularized in econometrics by *Sims (1980)*, which the main purpose of using it is to describe the evolution of a model's variables in reaction to a shock in one or more variables estimating standard deviation of the disturbance term to determine the impulse. This allows to trace the effects of an innovation shock to one variable on the response of all variables in the VAR model system for several periods in future, starting in t = 0 on the own variable to the others in t = 1 .

TABLE 3.8: Variance - covariance matrix VAR(1), m=3, "Source: own elaboration"

Variance - covariance matrix	**BTCCNY**	**BTCKRW**	**BTCJPY**
BTCCNY	0.001564549	0.001556504	0.001563982
BTCKRW	0.001556504	0.001569861	0.001561939
BTCJPY	0.001563982	0.001561939	0.001589296

In the previous table it can be seen that the off-diagonal elements of the estimated variance-covariance matrix are not zero, so it can be assumed that there is contemporaneous correlation between the variables in the VAR (1) model estimated.

The correlation matrix can be expressed as follows:

TABLE 3.9: Correlation matrix VAR(1), m=3, "Source: own elaboration"

Correlation matrix	**BTCCNY**	**BTCKRW**	**BTCJPY**
BTCCNY	1.0000000	0.9931735	0.9918244
BTCKRW	0.9931735	1.0000000	0.9888514
BTCJPY	0.9918244	0.9888514	1.0000000

But the main goal of using the **IR function** is to determine in which direction the impact will affect the behavior of the rest of the variables of the model, and the correlation matrix do not express this information at all.

An approach to identify how the shocks impact on the VAR(1) model estimated is computing **Orthogonalized Responses** using **Cholesky decomposition**.

The basic idea, well-developed in (Lütkepohl, 2006) is to decompose the previous estimated variance-covariance matrix to obtain the **lower triangular matrix** with positive diagonal elements through the use of Choleski decomposition. Doing this, it can be observed the sensitivity to a contemporaneous shock on each of the variables.

TABLE 3.10: OR - Choleski decomposition VAR(1), m=3, "Source: own elaboration"

Choleski decomposition	**BTCCNY**	**BTCKRW**	**BTCJPY**
BTCCNY	0.03955438	0.000000000	0.00000000
BTCKRW	0.03935100	0.004621719	0.00000000
BTCJPY	0.03954005	0.001297926	0.00491894

Note that the obtained matrix using **Choleski decomposition** is a lower triangular matrix, where the first equation does not have any other endogenous variables from the system, the second equation has the first two and the third has the first three.

Then, the first has only its own innovation, while the second has implicitly the first coming from the inclusion of the contemporaneous first variable. In other words, the variable in the first row (BTCCNY) will never be sensitive to a contemporaneous shock of any other variable, and the last variable (BTCJPY) in the system will be sensitive to shocks of all other variables. Graphically, the orthogonal impulse response from each of the variables has been moved to the **Appendix C**, but it can be deduced in the following points:

- The first plot shows the estimated standard deviation of the disturbance term as the response of BTCCNY to BTCKRW, which it has a minimum positive impact at $t = 1$ but then it gradually decreases until the effect of the shock disappears at $t = 2$.
- The second plot (Figure 2.5) shows the estimated standard deviation of the disturbance term as the response of BTCKRW to BTCCNY, which it has a no impact (not sensitive to a contemporaneous shock) at $t = 0$ but then it gradually increases at $t = 1$, where it hits its steady value, and then it decreases to the a value close to initial one at $t = 3$.

- The third plot (Figure 2.6) shows the estimated standard deviation of the disturbance term as the response of BTCJPY to BTCCNY, which it has a no impact at $t = 0$ but then it decreases at $t = 1$, where it hits its steady value, and then it increases to the a value close to initial one at $t = 3$.

3.2.3 Cointegration tests

It is normally assumed that when two or more series follow the same trend or pattern, that is, they tied together in the long term, these are said to be cointegrated.
Before start working on this concept, let's introduce some important rules of linear combinations of integrated series:

Single combination [x_{1t}]:

- If $x_{1,t} \sim$ I(0), then ω + ϕ $x_{1,t} \sim$ I(0)
- If $x_{1,t} \sim$ I(1), then ω + ϕ $x_{1,t} \sim$ I(1)

Linear combination [x_{1t}, x_{2t}]:

- If $x_{1,t} \sim$ I(0) and $x_{2,t} \sim$ I(0), then ω $x_{1,t}$ + ϕ $x_{2,t} \sim$ I(0)
- If $x_{1,t} \sim$ I(0) and $x_{2,t} \sim$ I(1), then ω $x_{1,t}$ + ϕ $x_{2,t} \sim$ I(1)
- If $x_{1,t} \sim$ I(1) and $x_{2,t} \sim$ I(1), then ω $x_{1,t}$ + ϕ $x_{2,t} \sim$ I(1)

But there is an exception that breaks this last rule and let to be ω x_{1t} **+** ϕ $x_{2t} \sim$ **I(0)**, and *Granger (1987)* pointed out that situation:

"A linear combination of two or more non stationary series can be stationary. If this linear stationary combination or I(0) exist, it is said that the series are cointegrated."

There is also one important case in which the previous statistical result is not applied, and it is when [x_{1t}, x_{2t}] $\sim$ I(1), but they are not cointegrated. This means that there is no linear combination of the variables that $\sim$ I(0). This situation was pointed out by *Granger and Newbold (1974)*, and it is known as the **spurious regression problem**.

It take place when is regressed one random walk onto another independent random walk and the estimated coefficient is significant, despite that the true value of the coefficient $\beta = 0$. So the residual is non stationary and the apparently significant relationship between two series just because they move together along the time are, in fact, unrelated series.

Mathematically, a linear combination using three time series (true related) that are independently non stationary, [x_{1t}, x_{2t}, x_{3t}], and grouped by a single vector X_t, such that:

$$Xt = [x_1 = (x_{1,1}, x_{1,2}, ..., x_{1,t}), x_2 = (x_{2,1}, x_{2,2}, ..., x_{2,t}), x_3 = (x_{3,1}, x_{3,2}, ..., x_{3,t})]$$

where they can be combined in a way that their linear combination is a stationary equilibrium relationship.

In the following expression, β represents what is known as the **cointegrating vector**:

$$\beta X_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} \sim I(0) \tag{3.18}$$

Because of the fact that there can be multiple cointegrating vectors that fit the same model, it has to be imposed a restriction to normalize the cointegrating vector for estimation, such as $\beta_1 = 1$.

$$\beta Y_t = y_{1,t} - \beta_2 y_{2,t} - \beta_3 y_{3,t} \sim I(0) \tag{3.19}$$
$$y_{1,t} = \beta_2 y_{2,t} + \beta_3 y_{3,t} + \mu_t,$$

By taking the residuals from a linear combination, such as:

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t \tag{3.20}$$
$$\hat{\epsilon}_t = y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t$$

If this $\hat{\epsilon} \sim I(0)$ are stationary, then x_t and y_t are cointegrated.

The *Engle and Granjer (1992)* cointegration test considers the case that there is a single cointegrating vector and a procedure method in two stages, $r = 1$ and $k = 2$, based on estimated residuals. It follows the the previous idea that if variables are cointegrated, then the residual of the cointegrated regression should be stationary.

$$x_{1,t} = \alpha + \beta 1 x_{2,t} + \epsilon_{1,t}, \tag{3.21}$$
$$x_{2,t} = \alpha + \beta 1 x_{1,t} + \epsilon_{2,t},$$

Regress and test for no-cointegration by testing for a unit root for each equations (ADF test):

$$\Delta\epsilon_{1,t} = \phi_1 \epsilon_{2,t-1} + \rho_{1,t}, \tag{3.22}$$
$$\Delta\epsilon_{2,t} = \phi_2 \epsilon_{2,t-1} + \rho_{2,t},$$

If it is not possible to reject the null hypotheses that $\phi_1, \phi_2 = 0$, then it cannot be rejected the hypothesis that this linear combination of variables are not cointegrated.

Johansen test

In this situation, it is gonna be applied the (Johansen, 1988) maximum eigenvalue and trace tests for cointegration under the empirically relevant situation of near-integrated variables. One of the main features in front of the Engle-Granger test is that this one permits more than one cointegrating relationship, so it can be applied more generally when working in situations with more than two variables or financial data, for example.

Trace tests and **Maximum eigenvalue tests** examine the number of linear combinations (r, rank) to be equal to a given value ($r*$) under the following statistic and hypothesis:

$$\lambda_{trace}(r) = -T \sum_{i=r*+1}^{n} ln(1-\hat{\lambda}_i) \tag{3.23}$$

$$\lambda_{max}(r, r+1) = -Tln(1-\hat{\lambda}_{r+1})$$

This means that for the trace test statistic, $\lambda_{trace}(r)$, it is contrasted the null hypothesis that the number of cointegrating vectors is $r = r* < k$ against the alternative hypothesis that the number of cointegrating vectors is $r = k$, where $k = p(lags) + 1$.

In the case of the maximum eigenvalue test statistic, $\lambda_{max}(r, r+1)$, it is contrasted the null hypothesis that the number of cointegrating vectors is $r = r* < k$ against the alternative hypothesis that the number of cointegrating vectors is $r = r * +1$.

TABLE 3.11: Trace test - VAR(p=1), k=2, "Source: own elaboration"

Rank	**Test**	**10 pct**	**5 pct**	**1 pct**
$r <= 2$	1.50	7.52	9.24	12.97
$r <= 1$	6.70	17.85	19.96	24.60
$r = 0$	37.06	32.00	34.91	41.07

TABLE 3.12: Maximum eigenvalue test - VAR(p=1), k=2, "Source: own elaboration"

Rank	Test	10 pct	5 pct	1 pct
$r <= 2$	1.50	7.52	9.24	12.97
$r <= 1$	5.20	13.75	15.67	20.20
$r = 0$	30.37	19.77	22.00	26.81

As it can be observed in the following tables, trace test identified at least one cointegrating vector while the maximum eigenvalue test found no cointegration for a model with a constant in the cointegrating equation. So it is has some conflict with cointegration rank tests, and that is because they test different alternative hypothesis, (Enders, 2014)

In this situation, it is more consistent (preferred) the output that reports the trace test, $(r = 1)$ in order to estimate the potential **Vector Error correction model** (VECM) and calculate its coefficients (β) as the long run cointegrating vector equation.

Cointegration equation (long-run model)

$$\begin{aligned} lnBTCCNY_t = & -105.4364 + 0.0056(lnBTCKRW_t) \\ & + 0.0037(lnBTCJPY_t) \end{aligned} \tag{3.24}$$

Estimated VECM with BTCCNY as target variable

$$\begin{aligned} \Delta lnBTCCNY_{t-1} = & -0.0588ECT_{t-1} - 0.7022\Delta(lnBTCCNY_{t-1}) + \\ & 0.0057\Delta(lnBTCKRW_{t-1}) - 0.0132\Delta(lnBTCJPY_{t-1}) \end{aligned} \tag{3.25}$$

Estimated VECM with BTCKRW as target variable

$$\begin{aligned} \Delta lnBTCKRW_{t-1} = & -7.6521ECT_{t-1} - 109.0729\Delta(lnBTCCNY_{t-1}) + \\ & 0.8747\Delta(lnBTCKRW_{t-1}) - 1.9267\Delta(lnBTCJPY_{t-1}) \end{aligned} \tag{3.26}$$

Estimated VECM with BTCJPY as target variable

$$\begin{aligned} \Delta lnBTCJPY_{t-1} = & -0.4829ECT_{t-1} - 8.5105\Delta(lnBTCCNY_{t-1}) + \\ & 0.0852\Delta(lnBTCKRW_{t-1}) - 0.3006\Delta(lnBTCJPY_{t-1}) \end{aligned} \tag{3.27}$$

In the **Appendix C** there are attached three tables of the R outputs obtained with the significance level and any other important information for each of the short-run equations [$\Delta BTCCNY_t$, $\Delta BTCKRW_t$, $\Delta BTCJPY_t$] and **Appendix D** contains the plotted forecast of the same model responses.

ARDL Bounds Testing approach

The non-derivation of a possible consistent solution to our financial series, and especially when it has been trying to demonstrate the existence of a correlation between them, has led to search for alternative literature about the models that are currently trying to measure the existence of cointegration. It is taken for granted that the series that have been exposed in this work show a strong pattern of movement and long-term trend that clearly demonstrate this phenomenon, but for some reason derived from the analysis of the variables, it has not been possible to reach a definitive solution even assuring the requirements for the application of the Johansen's methodology.

The **Autoregressive Distributed Lags** (ARDL) cointegration methodology, (Pesaran and Shin, 1999), is based on determining the long run relationship between series following Eagle and Granger method of cointegration with stationary series, integrated of the same order and cointegrated. The model contains the lagged value of the dependent variable, the current value and the lagged value of regressors as explanatory variables, combining endogenous and exogenous variables - unlike a VAR model that is strictly for endogenous variables.

ARDL test equation:

$$\begin{aligned} ARDL(p,q) : yt = y_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} \\ + \gamma_1 x_t + \gamma_2 x_{t-1} + ... + \gamma_q x_{t-q} + \epsilon_t \end{aligned} \tag{3.28}$$

Then, summarizing lagged values of y and x (values of the regressors and its lagged values) it is obtained the following equation:

$$ARDL(p,q) : yt = y_0 + \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{i=0}^{q} \gamma_i x_{t-i} + \epsilon_t$$

where the variables in x_t are allowed to be I(0), I(1) or cointegrated variables. The **ARDL bound test**, (Pesaran and Smith, 2001), is based on determining the long run relationship between series with **different** order of integration - independent variables are allowed to be individually either I(0) or I(1) but lower than I(2), to obtain the short-run dynamics and long-run relationship.

ARDL bound test equation:

$$ARDL : \Delta yt = y_0 + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + \sum_{i=0}^{q} \gamma_i \Delta x_{t-i} + \phi_1 y_{t-1} + \phi_2 x_{t-1} \epsilon_t,$$

where the new adjusted model, respect the previous one, contemplates the components of:

- Short-run (ρ_i,γ_i) such as: $\sum_{i=1}^{p} \rho_i \Delta y_{t-i} + \sum_{i=0}^{q} \gamma_i \Delta x_{t-i}$.
- Long-run (ϕ_1, ϕ_2) such as: $\phi_1 y_{t-1} + \phi_2 x_{t-1}$.
- Error/disturbance such as: ϵ_t.

The steps to follow to develop the estimation process of this adjusted model by Pesaran is structured quite similar as the Johansen method, but with certain alterations in the structure.

1. Identify serial correlation.
2. If there are serially correlated errors, then Least Square Estimation - **HAC standard errors**.
3. Estimate an **AR(p) Error Model**, Non linear Least Squares Estimation.
4. Determinate **ARDL bound test model**.
5. Determinate the optimal lag structure.
6. Apply cointegration test on the series.
7. Estimate **long-run levels model** to obtain **ECT** (residuals).
8. Estimate **VECM** with the estimated residuals.
9. Estimate **short-run model** (short-run dynamics).

From our data, to perform the ARDL bound test for cointegration, the conditional model with three variables is specified as:

$$\begin{aligned} \Delta BTCCNY_t = \alpha_0 + \sum_{i=1}^{p} \rho_i \Delta BTCCNY_{t-i} + \sum_{i=0}^{q1} \gamma_i \Delta BTCKRW_{t-i} + \\ \sum_{i=0}^{q2} \omega_i \Delta BTCJPY_{t-i} + \phi_0 BTCCNY_{t-1} + \\ \phi_1 BTCKRW_{t-1} + \phi_2 BTCJPY_{t-1} + \epsilon_t \end{aligned} \tag{3.29}$$

Under the **F-bound test** null and alternative hypothesis:

$$\begin{aligned} H_0 &: \phi_0 = \phi_1 = \phi_2 = 0 \\ H_1 &: \phi_0 \neq \phi_1 \neq \phi_2 \end{aligned} \tag{3.30}$$

where the null hypothesis exposes that "a long-run relationship does not exists" and the alternative hypothesis that "a long-run relationship exists".

TABLE 3.13: Bounds F-test (Wald), "Source: own elaboration"

statistic	Lower-bound I(0)	Upper-bound I(1)	p-value
-3.8422	-2.86	-3.53	0.02142

According to Pesaran (2001), there are 5 differents cases for testing the cointegration bound test. In our case, it is considered the case 3 where there **is/are unrestricted/s intercept/s and no trends**. As it is shown in the table, the Wald statistic (-3.84) falls above the upper-bound (-3.53) and then it can be safely said that **exists cointegration**. If it falls between the lower-bound and the upper-bound critical value, then it is considered inconclusive. And if it falls below the lower-bound, then it is said that there is no cointegration.

All of this under the selected ARDL model (4,5,5) - (limited to a max order of 5 to avoid overparameterization):

TABLE 3.14: ARDL model selection, "Source: own elaboration"

BTCCNY	BTCKRW	BTCJPY	AIC
4	5	5	21651.76
5	5	5	21653.56
3	5	1	21657.27
2	5	1	21658.98
3	5	2	21659.05
...	...	...	...
2	2	2	21697.05

In the **Appendix E** there are attached three tables of the R outputs obtained with the significance level and any other important information for each of the coefficients of BTCCNY, BTCKRW, BTCJPY and their different orders.

Additionally, it is applied the **Durbin-Watson test** to ensure that there is no autocorrelation of residuals on this linear regression fit, (Durbin and Watson, 1950; Durbin and Watson, 1951). The obtained output shows that:

TABLE 3.15: Durbin-Watson test, "Source: own elaboration"

Durbin-Watson statistic	p-value
1.9888	0.7963

Then, it cannot be rejected the null hypothesis that there is **no autocorrelation of residuals**. The same criteria can be applied to ensure the existence of homoscedasticity (variance does not depend on auxiliary regressors) using the studentized Breusch-Pagan test, (Breusch and Pagan, 1979), as it follows:

TABLE 3.16: Breusch-Pagan test, "Source: own elaboration"

Breusch-Pagan statistic	**p-value**
78.962	0.0821

In this case the p-value is slightly higher than 0.05, but again it cannot be rejected the null hypothesis of the existence of **homoscedasticity**.

In order to determine that the model is well-specified and there is no misspecification functional form, such as the model does not account for some important nonlinearities or omit important variables, it can be applied the **Ramsey RESET test**, introduced by (Ramsey, 1969).

TABLE 3.17: Ramsey RESET test, "Source: own elaboration"

Ramsey RESET statistic	**p-value**
0.045941	0.9551

Finally, it is estimated the **Restricted Error Correction Model (RECM)** under the case 3 (unrestricted intercepts and no trend, $c_1 = 0$), where applying our data it is obtained the following equations:

$$\begin{aligned}\Delta lnBTCCNY_t = \alpha_0 + \sum_{i=1}^{3} \rho_i \Delta lnBTCCNY_{t-i} + \sum_{i=0}^{4} \gamma_i \Delta lnBTCKRW_{t-i} + \\ \sum_{i=0}^{4} \omega_i \Delta lnBTCJPY_{t-i} + \lambda_3 ECT_{t-1} + \epsilon_t\end{aligned} \tag{3.31}$$

where this equation represents the **short-run dynamic model** and the coefficient of the parameter error correction term, ECT_{t-1}, represents the speed of adjustment towards equilibrium. In the **Appendix F** there are attached R outputs with the significance level and important information for each of the parameters.

The **long-run model** can be written taking into consider the next multipliers (elasticities):

$$lnBTCCNY_t = -115.1117 + 0.005564(lnBTCKRW_t) + 0.004261(lnBTCJPY_t) \tag{3.32}$$

Chapter 4

Conclusions and research limitations

4.1 Conclusions

The whole process to get to the formulation of the cointegration model has been adjusted well during the process. It has been shown that the temporary series of the prices are non stationary and integrated of order 1, I(1), something needed to evaluate our goal. Parallelly, it has been observed that the returns of each of the variables are stationary and integrated of order 0, I(0), something predictable after looking at the series of returns in the first figure in the previous chapter. Once reviewed, it has been proposed a VAR(1) after determining the optimal number of lags to use to compute the **Johansen test**.

Then it has been evaluated the **Granger causality** test, where with a VAR(1) it has not been observed any variable that was Granger-cause to the other ones. Strictly speaking, one variable is Granger causal for another one if the first one helps predict the second one at some stage in the future and non of the selected variables seem to be showing this behaviour.

In the same way, it has been observed that the **Impulse response function** (IRF) tracks the impact of any variable on others in the system but as a very short term impact, because after period 2 or 3 (depending on the variable) the response is not significantly different from zero (at the 5% level) and the idea of a % change in something of the first variable produces a % change in something of the second variable and its effect disappears.

Finally, according to the literature and the reasoning established by the authors, if the long-run equilibrium shows the relationship between the variables **without any short-run shock** or the relationship from which variables deviate but always return to, then it can be shown that the estimated long-run model equation seems to be accurate. Applying the

Johansen test, it has been derived an equation to explain the equilibrium in the future, but the results of the coefficients, especially the error-correction coefficient for every dynamic short-run model estimated, does not seem to be statistically significant at any confidence level.

For example, if it is selected the dynamic short-run model for the target variable BTC-KRW, it is observed that the value of this error-correction coefficient is **-7.6521**, which is too high to be a value delimited between [-1,0]. This interval has to be imposed because this coefficient captures the speed of adjustment such that when the dependent variable exceeds the long-run relationship, this has to be corrected to revert to the steady state. Considering this value, this coefficient means that about **765.25** % of this disequilibrium is corrected between 1 year (annualized data) - and this does not make sense.
A plausible argument that can justify this high value is probably explained by the fact that prices are national prices, not standardized, and therefore, there is a large difference in volume in them. So the correction factor may need to be scaled to higher volumes as it is not standardized to a single currency.

From the cointegration equation (long-run) , it has been shown that an increment of 1% of the prices of BTC-CNY increases the prices of BTC-KRW in about **+0.56**%, but what happens to BTC-JPY, **+0.37**%, seems to be not significant in order to get this equilibrium in the future. These values seem to be quite accurate, but as it has been said before, what differs in the integrity of the estimated model are the coefficient values obtained in dynamic short-run equations and increasing the lag order established in the VAR model is not an option, since it has been tested and has not turned out to be the solution.

Because of that, an **Autoregressive Distributed Lag (ARDL) bounds testing approach** has been proposed to determine the existence of cointegration and thus be able to contrast the results obtained with the Johansen test. The selected model, an ARDL (4,5,5), has passed the **F-bound test** to prove the existance of long-run relationship and it has been ensured that there is no autocorrelation of residuals applying the use of the **Durbin-Watson test**. At the same time, it has been proved using the **Breusch-Pagan test** that the variables have the same finite variance (homoscedasticity). Finally, with the **Ramsey RESET test**, it has been determined that the structure of the model has as an adequate functional form.

From the cointegration equation (long-run), it has been shown that an increment of 1% of the prices of BTC-CNY increases the prices of BTC-KRW in about **+0.56**%, and, unlike what it has obtained with the Johansen test, what happens to BTC-JPY, **+0.43**%, seems to be significant in order to get this equilibrium in the future.

These values are quite similar in both resultant equations, and this allows us to ensure that there is a **long-term relationship** of Bitcoin series and that the dynamic adjustment equations have a significant impact when determining the steady state in each of the observed variables.

4.2 Limitations of the study

The results obtained are subject to the limitations existing in the applied methodology and the following literature gaps that has been identified. Considering the Johansen test to obtain the cointegration equation and its respective dynamics adjustment, it is known that is an improvement over the "initial" cointegration test exposed by Engle-Granger test, as (Bilgili, 1998; Gonzalo and Lee, 1998; Haug, 1996) point out in their respective research. It can detect multiple cointegrating vectors and it is more consistent for multivariate analysis, which is one of the objectives of the study. Another robust property is that Johansen test deals with the variables as endogenous variables once it is computed the model, as it has been observed when determining the dynamic adjustment equations for each one of the analyzed variables (Johansen, 1995).

But, as it has been reported by (Gonzalo and Lee, 1998), in some scenarios Engle-Granger test can be more robust than Johansen likelihood ratio test. This can be pointed out in some research papers that compare performance of both tests, (Wee and Tan, 1997; Agoraki and Kouretas, 2019; Pascual, 2003), and this is why it has been selected Johansen as a main methodology to follow and compare with alternative models to study the cointegration effect than Engle-Granger, where in that case it is selected an ARDL Bounds Testing approach.

One of the arguments that had to be justified has been the high values obtained in the dynamic equations of BTC-KRW, specially the speed of adjustment: **(-7.6521)**, and why the coefficients obtained are not significant at any confidence level, except two or three. This situation is observed and suggested in other levels by different authors, such as (Maggiora and Skerman, 2009).

In that case, once the ARDL bounds testing approach model is computed, it is obtained that most of the parameters are significant and in the case of BTC-KRW they are very much lower than obtained with the Johansen methodology. The main empirical argument that justifies this situation is marked by the ARDL model being less restrictive to raise the order of each of the variables to adequate levels to adjust the dynamic adjustment equation.

This is observed by (Haq and Larsson, 2016; Musa and Zoramawa, 2014; Bouri and Shahbaz, 2018) in their studies and the methodology used to diagnose the value of the parameters obtained in their series. Another reason can be explained, as it has been exposed before in the conclusions section, by the fact that prices are not standardized, and therefore, there is a large difference in volume. This makes that the correction factor may need to be scaled to higher volumes as it is not standardized to a single currency (Biggeri and Ferrari, 2012).

Another limitation or hypothesis adopted in the formulation of the ARDL model is that it is limited to a maximum order of 5. This ensures that there is no excess of parameters in the model, but also restricts that for values greater than 5 it can be found a model that better fits our data. Many authors agree that this situation may effect the results and the optimal model selection, (Wang and Cao, 2020; Tursoy, 2019), but in our data it seems that the selected ARDL (4,5,5) model order fits quite well.

Finally, the time period chosen for the analysis of the time series has coincided with a period of price increased and optimal economic situation. The impact derived from COVID-19 could affect the projections that can be made based on this established model or the cointegration equation and the long-term equilibrium for projections beyond this year, which would cause the model and the calculations made with the Johansen test had to be calibrated again. Many authors have already tried to capture the impact derived from the pandemic and have established correction hypotheses using ARDL models to quantify the impact, for example, in crude oil prices or tourism, (Albulescu, 2020; Ghosh, 2020).

But even so, the goals of the thesis have been met, opening the line for a further research with alternative approaches such as the use of machine learning algorithms or bayesian neuronal models.

Appendix A

Additional ADF, PP and KPSS tables

TABLE A.1: ADF, PP and KPSS tests of BTC-KRW series (prices), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	-1.173	-1.1915	4.79
T-statistic value	-2.86	-2.863615	0.463
Results	NRH0	NRH0	RH0

TABLE A.2: ADF, PP and KPSS tests of BTC-KRW series (returns), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	-30.6073	-42.5698	0.1653
T-statistic value	-2.86	-2.863616	0.463
Results	RH0	RH0	NRH0

TABLE A.3: ADF, PP and KPSS tests of BTC-JPY series (prices), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	-1.3436	-1.4288	4.6408
T-statistic value	-2.86	-2.863615	0.463
Results	NRH0	NRH0	RH0

TABLE A.4: ADF, PP and KPSS tests of BTC-JPY series (returns), "Source: own elaboration"

Test	ADF	**PP**	**KPSS**
Null hypothesis	Non stationary	Non stationary	Stationary
P-value/T-value	-30.6235	-42.718	0.1677
T-statistic value	-2.86	-2.863616	0.463
Results	RH0	RH0	NRH0

Appendix B

Orthogonal Impulse Response

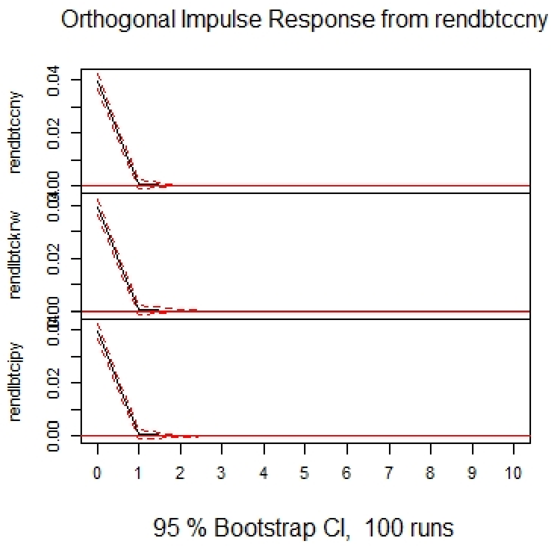

FIGURE B.1: Orthogonal Impulse Response from BTCCNY, "Source: own elaboration"

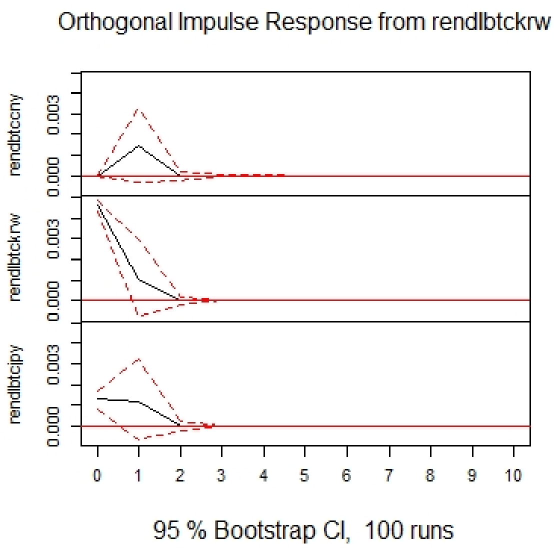

FIGURE B.2: Orthogonal Impulse Response from BTCKRW, "Source: own elaboration"

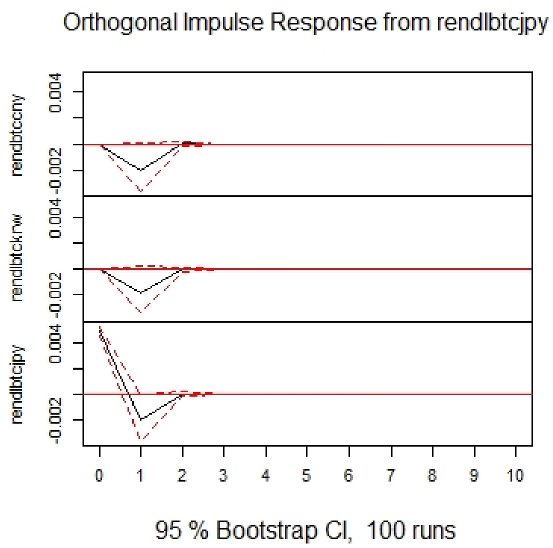

FIGURE B.3: Orthogonal Impulse Response from BTCJPY, "Source: own elaboration"

Appendix C

Short-run response coefficients

```
Response btccny.d :

Call:
lm(formula = btccny.d ~ ect1 + btccny.dl1 + btckrw.dl1 + btcjpy.dl1 -
    1, data = data.mat)

Residuals:
     Min       1Q   Median       3Q      Max
-15680.1    -79.8      6.7    178.7  23054.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
ect1       -0.058764   0.070240  -0.837  0.40292
btccny.dl1 -0.702191   0.448343  -1.566  0.11748
btckrw.dl1  0.005753   0.002169   2.652  0.00806 **
btcjpy.dl1 -0.013198   0.015200  -0.868  0.38535
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1861 on 1805 degrees of freedom
Multiple R-squared:  0.005899,  Adjusted R-squared:  0.003696
F-statistic: 2.678 on 4 and 1805 DF,  p-value: 0.03033
```

FIGURE C.1: Response BTCCNY coefficients,
“Source: own elaboration”

```
Response btckrw.d :

Call:
lm(formula = btckrw.d ~ ect1 + btccny.dl1 + btckrw.dl1 + btcjpy.dl1 -
    1, data = data.mat)

Residuals:
     Min       1Q   Median       3Q      Max
-2580883   -13393     1395    30062  3820067

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
ect1         -7.6521    11.7155  -0.653   0.5137
btccny.dl1 -109.0729    74.7801  -1.459   0.1449
btckrw.dl1    0.8747     0.3618   2.418   0.0157 *
btcjpy.dl1   -1.9267     2.5353  -0.760   0.4474
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 310500 on 1805 degrees of freedom
Multiple R-squared:  0.004457,  Adjusted R-squared:  0.002251
F-statistic:  2.02 on 4 and 1805 DF,  p-value: 0.08916
```

FIGURE C.2: Response BTCKRW coefficients,
“Source: own elaboration”

```
Response btcjpy.d :

Call:
lm(formula = btcjpy.d ~ ect1 + btccny.dl1 + btckrw.dl1 + btcjpy.dl1 -
    1, data = data.mat)

Residuals:
    Min      1Q  Median      3Q     Max
-271845   -1336     153    2977  406043

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
ect1       -0.48290    1.17670  -0.410   0.6816
btccny.dl1 -8.51047    7.51089  -1.133   0.2573
btckrw.dl1  0.08519    0.03634   2.344   0.0192 *
btcjpy.dl1 -0.30058    0.25464  -1.180   0.2380
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31180 on 1805 degrees of freedom
Multiple R-squared:  0.005468,	Adjusted R-squared:  0.003265
F-statistic: 2.481 on 4 and 1805 DF,  p-value: 0.04209
```

FIGURE C.3: Response BTCJPY coefficients,
"Source: own elaboration"

Appendix D

Plotting VECM model responses

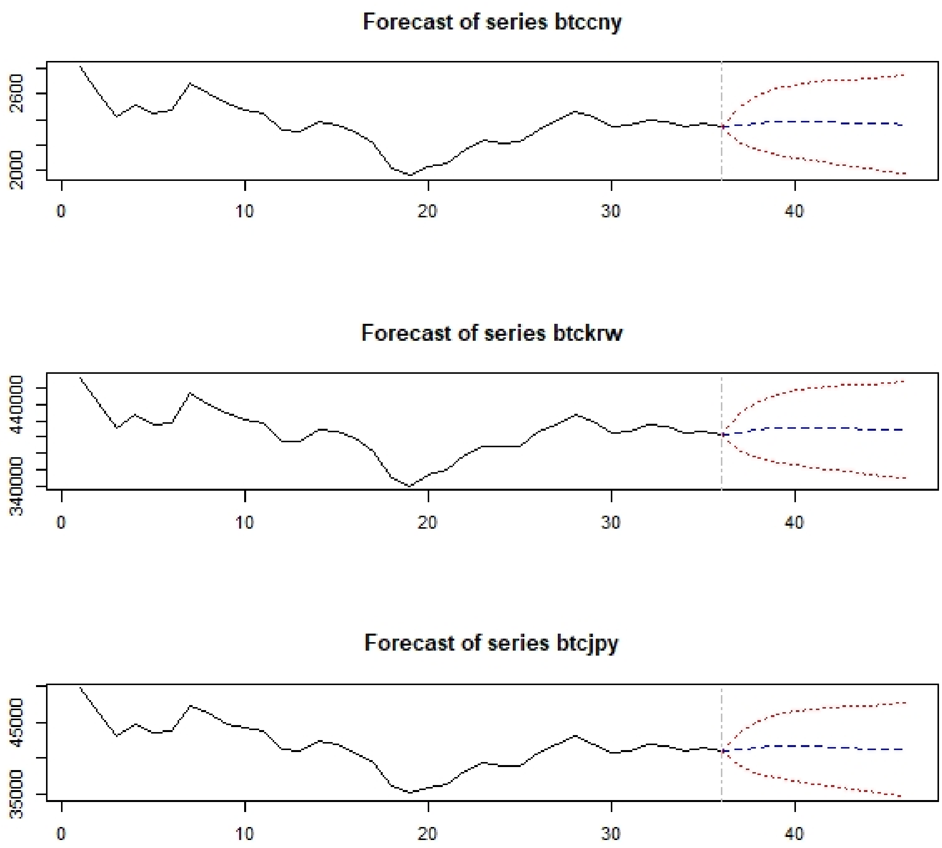

FIGURE D.1: VECM model respones, "Source: own elaboration"

Appendix E

ARDL model coefficients

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.105e+00  3.085e+00  -0.682 0.495187
L(btccny, 1)    -1.829e-02  4.759e-03  -3.842 0.000126 ***
L(btckrw, 1)     1.018e-04  2.262e-05   4.498 7.29e-06 ***
L(btcjpy, 1)     7.791e-05  1.040e-04   0.749 0.453800
d(L(btccny, 1)) -8.145e-02  2.359e-02  -3.452 0.000568 ***
d(L(btccny, 2))  4.248e-02  2.372e-02   1.791 0.073477 .
d(L(btccny, 3))  4.460e-02  2.370e-02   1.882 0.059982 .
d(btckrw)        4.003e-03  6.459e-05  61.979  < 2e-16 ***
d(L(btckrw, 1))  4.757e-04  1.151e-04   4.134 3.73e-05 ***
d(L(btckrw, 2)) -2.092e-04  1.156e-04  -1.809 0.070571 .
d(L(btckrw, 3)) -8.774e-05  1.158e-04  -0.757 0.448950
d(L(btckrw, 4))  2.211e-04  6.514e-05   3.394 0.000705 ***
d(btcjpy)        1.982e-02  6.430e-04  30.828  < 2e-16 ***
d(L(btcjpy, 1))  4.139e-04  8.000e-04   0.517 0.605000
d(L(btcjpy, 2)) -3.958e-04  8.022e-04  -0.493 0.621771
d(L(btcjpy, 3)) -1.716e-03  7.984e-04  -2.149 0.031770 *
d(L(btcjpy, 4)) -2.031e-03  6.469e-04  -3.139 0.001721 **
```

FIGURE E.1: ARDL model coefficients,
"Source: own elaboration"

Appendix F

RECM coefficient values

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.105e+00	2.337e+00	-0.901	0.367914	
d(L(btccny, 1))	-8.145e-02	2.341e-02	-3.480	0.000514	***
d(L(btccny, 2))	4.248e-02	2.351e-02	1.806	0.071009	.
d(L(btccny, 3))	4.460e-02	2.353e-02	1.896	0.058155	.
d(btckrw)	4.003e-03	6.450e-05	62.061	< 2e-16	***
d(L(btckrw, 1))	4.757e-04	1.147e-04	4.148	3.51e-05	***
d(L(btckrw, 2))	-2.092e-04	1.152e-04	-1.816	0.069516	.
d(L(btckrw, 3))	-8.774e-05	1.155e-04	-0.760	0.447384	
d(L(btckrw, 4))	2.211e-04	6.504e-05	3.399	0.000692	***
d(btcjpy)	1.982e-02	6.421e-04	30.872	< 2e-16	***
d(L(btcjpy, 1))	4.139e-04	7.950e-04	0.521	0.602732	
d(L(btcjpy, 2))	-3.958e-04	7.969e-04	-0.497	0.619436	
d(L(btcjpy, 3))	-1.716e-03	7.943e-04	-2.160	0.030903	*
d(L(btcjpy, 4))	-2.031e-03	6.461e-04	-3.143	0.001697	**
ect	-1.829e-02	3.917e-03	-4.668	3.26e-06	***

FIGURE F.1: Short-run dynamic and long-run coefficients, "Source: own elaboration"

Appendix G

ARDL - Cointegration equation

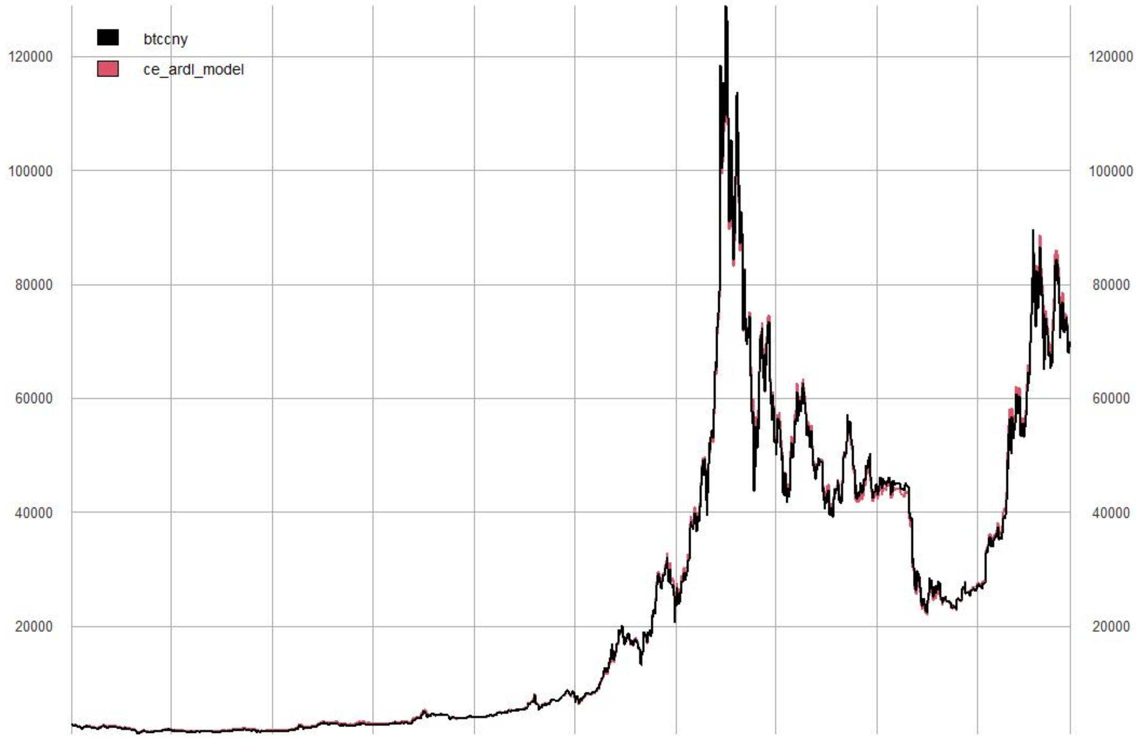

FIGURE G.1: Cointegration equation - Projected vs Observed, "Source: own elaboration"

Bibliography

Agoraki M-E., Georgoutsos D. and G. Kouretas (2019). "Capital markets integration and cointegration: Testing for the correct specification of stock market indices." In: *Journal of Risk and Financial Management* 12.

Akaike, H. (1974). "A new look at the statistical model identification." In: *IEEE Transactions on Automatic Control* AC-19, pp. 716–723.

Albulescu, C. (2020). "Do COVID-19 and crude oil prices drive the US economic policy uncertainty?." In: *HAL Id: hal-02509450*. URL: https://hal.archives-ouvertes.fr/hal-02509450.

Biggeri, L. and G. Ferrari (2012). "Price indexes in time and space: methods and practice." In: *Physica* Contributions to Statistics.

Bilgili, F. (1998). "Stationarity and cointegration tests: Comparison of Engle - Granger and Johansen methodologies." In: *Journal of Faculty of Economics and Administrative Sciences, Erciyes University* 13, pp. 131–141.

Bouri E., Grupta R. Lahiani A. and M. Shahbaz (2018). "Testing for asymmetric nonlinear short- and long-run relationships between bitcoin, aggregate commodity and gold prices." In: *Resources Policy* 57, pp. 224–235.

Breusch, T. and R. Pagan (1979). "A simple test for heteroscedasticity and random coefficient variation." In: *Econometrica* 47, 1287–1294.

Dickey, D. A. and W. A. Fuller (1979). "Distributions of the estimators for autoregressive time series with a unit root." In: *Journal of the American Statistical Association.* 75, pp. 427–431.

Durbin, J. and G. Watson (1950). "Testing for serial correlation in least squares regression I." In: *Biometrika* 37, pp. 409–428.

— (1951). "Testing for serial correlation in least squares regression II." In: *Biometrika* 38, pp. 159–178.

Edgerton, D. and G. Shukur (1999). "Testing autocorrelation in a system perspective." In: *Econometric Reviews* 18, pp. 43–386.

Enders, W. (2014). "Applied Econometric Time Series". In: *Wiley* 3.

Engle, R.F. and C. W. J. Granjer (1987). "Co-integration and error correction: Representation, estimation and testing." In: *Econometrica* 55, 251–276.

Fry, J. and E. T. Cheah (2016). "Negative bubbles and shocks in cryptocurrency markets." In: *International Review of Financial Analysis* 47, pp. 343–352.

Ghosh, S. (2020). “Asymmetric impact of COVID-19 induced uncertainty on inbound Chinese tourists in Australia: insights from nonlinear ARDL model.” In: *Quantitative Finance and Economics* 4, pp. 343–364.

Gonzalo, J. and T.H. Lee (1998). “Pitfalls in testing for long run relationships.” In: *Journal of Econometrics* 86, pp. 129–154.

Granger, C. W. J. (1969). “Investigating causal relations by econometric models and cross-spectral methods.” In: *Econometrica* 37, pp. 424–438.

Hamilton, J. (1994). “Time series analysis.” In: *Princeton University Press* 0.

Haq, S. and R. Larsson (2016). “The dynamics of stock market returns and macroeconomic indicators: An ARDL approach with cointegration.” In: *Master of Science Thesis INDEK 2016:59* KTH Industrial Engineering and Management, Stockholm.

Haug (1996). “Tests for cointegration A Monte Carlo comparison.” In: *Journal of Econometrics* 71, pp. 89–115.

Johansen, S. (1988). “Statistical analysis of cointegration vectors.” In: *Journal of Economic Dynamics and Control.* 12, pp. 231–254.

— (1995). “Identifying restrictions of linear equations with applications to simultaneous equations and cointegration.” In: *Journal of Econometrics* 69, 111–132.

Kwiatkowski D., Phillips P. C. B. Schmidt P. and Y. Shin (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?.” In: *Journal of Econometrics* 54, pp. 159–178.

Ljung, G.M. and G.E.P. Box (1978). “On a measure of lack of fit in time series models.” In: *Biometrika* 65, pp. 297–303.

Lütkepohl, H. (2006). “New Introduction to Multiple Time Series Analysis.” In: *Springer*.

Maggiora, D. and R. Skerman (2009). “Johansen cointegration analysis of american and european stock market indices.” In: *Master Thesis* Lund University.

Musa Y., Usman U. and A. Zoramawa (2014). “Relationship between money supply and government revenues in Nigeria.” In: *CBN Journal of Applied Statistics* 5, pp. 117–136.

Narayanan A., Bonneau J. Felten E. Miller A. and S. Goldfeder (2016). “Bitcoin and cryptocurrency technologies: a comprehensive introduction.” In: *Princeton University Press.* 336.

Pascual, A. (2003). “Assessing European stock markets (co)integration.” In: *Economics Letters* 78, 197–203.

Pesaran, M. and Y. Shin (1999). “An Autoregressive Distributed-Lag Modelling Approach to Cointegration Analysis.” In: *Cambridge University Press* Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium.371-413.

Pesaran M., Shin Y. and R. Smith (2001). "Bounds testing approaches to the analysis of level relationships." In: *Journal of Applied Econometrics* 16.289-326.

Phillips, P.C.B. and P. Perron (1988). "Testing for a unit root in time series regression." In: *Biometrika* 75(2), pp. 335–346.

Quinn, B. (1980). "Order determination for a multivariate autoregression." In: *Journal of the Royal Statistical Society* B42, pp. 182–185.

Ramsey, J. (1969). "Tests for specification error in classical linear least squares regression analysis." In: *Journal of the Royal Statistical Society* 31, 350–371.

Schwarz, G. (1978). "Estimating the dimension of a model." In: *Annals of Statistics* 6, pp. 461–464.

Tursoy, T. (2019). "The interaction between stock prices and interest rates in Turkey: empirical evidence from ARDL bounds test cointegration." In: *Financial Innovation* 5.

Wang Q., Hao Y. and J. Cao (2020). "ADRL: An attention-based deep reinforcement learning framework for knowledge graph reasoning." In: *Knowledge-Based Systems* 197.

Wee, C. and R. Tan (1997). "Performance of Johansen's cointegration test." In: *East Asian Economic Issues*, pp. 402–414.

Zivot, E. and Donald W.K. Andrews (1992). "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." In: *Journal of Business and Economic Statistics* 10(3), pp. 251–270.

```
###########Packages and libraries###########

install.packages("car")
install.packages("urca")
install.packages("vars")
install.packages("tseries")
install.packages("ggplot2")
install.packages("tsDyn")
install.packages("forecast")
install.packages("fGarch")
install.packages("quantmod")
library(car)
library(urca)
library(vars)
library(tseries)
library(ggplot2)
library(tsDyn)
library(forecast)
library(fGarch)
library(quantmod)
library(ARDL)

###########Data###########

getSymbols("BTC-CNY",from="2014-09-01", to="2019-09-01", periodicity="daily")
getSymbols("BTC-KRW",from="2014-09-01", to="2019-09-01", periodicity="daily")
getSymbols("BTC-JPY",from="2014-09-01", to="2019-09-01", periodicity="daily")

#Price selection - adjusted prices and format dates#
```

```
btccny=`BTC-CNY`[,6]
btckrw=`BTC-KRW`[,6]
btcjpy=`BTC-JPY`[,6]
Dates<-as.Date(rownames(zoo(`BTC-CNY`)))
#Dates<-as.Date(rownames(zoo(`BTC-KRW`)))# Same length on each case.
#Dates<-as.Date(rownames(zoo(`BTC-JPY`)))# Same length on each case.
btccny<-as.numeric(`BTC-CNY`$`BTC-CNY.Adjusted`)
btckrw<-as.numeric(`BTC-KRW`$`BTC-KRW.Adjusted`)
btcjpy<-as.numeric(`BTC-JPY`$`BTC-JPY.Adjusted`)

#Derive returns each series and alter [-1] date structure#

rendbtccny <- diff(log(btccny))
rendbtckrw <- diff(log(btckrw))
rendbtcjpy <- diff(log(btcjpy))
diffDates<-as.Date(Dates[-1],"%d/%m/%Y")

#basicStats() function to statistic summary#

basicStats(btccny)
basicStats(btckrw)
basicStats(btcjpy)
basicStats(rendbtccny)
basicStats(rendbtckrw)
basicStats(rendbtcjpy)

#Plot grouped data series#

win.graph(width=8,height=5)
par(mfrow=c(3,2),font=3,font.lab=4,font.axis=2,las=1)
plot(Dates,btccny,type="l",col="black",main="BTC - CNY Prices", ann=FALSE)
```

```
plot(diffDates,rendbtccny,type="l",col="black",main="BTC - CNY Returns", ann=FALSE)
plot(Dates,btckrw,type="l",col="black",main="BTC - KRW Prices", ann=FALSE)
plot(diffDates,rendbtckrw,type="l",col="black",main="BTC - KRW Returns", ann=FALSE)
plot(Dates,btcjpy,type="l",col="black",main="BTC - JPY Prices", ann=FALSE)
plot(diffDates,rendbtcjpy,type="l",col="black",main="BTC - JPY Returns", ann=FALSE)

##########Stantionary analysis of the series - Informal##########

#Grouped ACF-PACF functions#

win.graph(width=8,height=5)
par(mfrow=c(2,3),font=2,font.lab=4,font.axis=2,las=1)
acf(btccny,ylim=c(-1,1),main="BTC-CNY")
pacf(btccny,ylim=c(-1,1),main="BTC-CNY")
acf(btckrw,ylim=c(-1,1),main="BTC-KRW")
pacf(btckrw,ylim=c(-1,1),main="BTC-KRW")
acf(btcjpy,ylim=c(-1,1),main="BTC-JPY")
pacf(btcjpy,ylim=c(-1,1),main="BTC-JPY")
win.graph(width=8,height=5)
par(mfrow=c(2,3),font=2,font.lab=4,font.axis=2,las=1)
acf(rendbtccny,ylim=c(-1,1),main="BTC-CNY")
pacf(rendbtccny,ylim=c(-1,1),main="BTC-CNY")
acf(rendbtckrw,ylim=c(-1,1),main="BTC-KRW")
pacf(rendbtckrw,ylim=c(-1,1),main="BTC-KRW")
acf(rendbtcjpy,ylim=c(-1,1),main="BTC-JPY")
pacf(rendbtcjpy,ylim=c(-1,1),main="BTC-JPY")

#Box-Pierce and Ljung-Box tests#

Box.test(btccny, lag = 1, type = c("Ljung-Box")) #replace for returns series
Box.test(btccny, lag = 5, type = c("Ljung-Box"))
```

```
Box.test(btccny, lag = 10, type = c("Ljung-Box"))
Box.test(btccny, lag = 15, type = c("Ljung-Box"))
Box.test(btccny, lag = 20, type = c("Ljung-Box"))
Box.test(btckrw, lag = 1, type = c("Ljung-Box"))
Box.test(btckrw, lag = 5, type = c("Ljung-Box"))
Box.test(btckrw, lag = 10, type = c("Ljung-Box"))
Box.test(btckrw, lag = 15, type = c("Ljung-Box"))
Box.test(btckrw, lag = 20, type = c("Ljung-Box"))
Box.test(btcjpy, lag = 1, type = c("Ljung-Box"))
Box.test(btcjpy, lag = 5, type = c("Ljung-Box"))
Box.test(btcjpy, lag = 10, type = c("Ljung-Box"))
Box.test(btcjpy, lag = 15, type = c("Ljung-Box"))
Box.test(btcjpy, lag = 20, type = c("Ljung-Box"))

###########Stantionary analysis of the series - Formal###########

###Prices series###

#Augmented Dickey-Fuller test#

btccny.df<-ur.df(btccny, type = c("drift"), selectlags = c("BIC"))
summary(btccny.df)
btckrw.df<-ur.df(btckrw, type = c("drift"), selectlags = c("BIC"))
summary(btckrw.df)
btcjpy.df<-ur.df(btcjpy, type = c("drift"), selectlags = c("BIC"))
summary(btcjpy.df)

#Phillips-Perron test#

btccny.pp<-ur.pp(btccny, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(btccny.pp)
```

```
btckrw.pp<-ur.pp(btckrw, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(btckrw.pp)
btcjpy.pp<-ur.pp(btcjpy, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(btcjpy.pp)

#Kwiatkowski-Phillips-Schmidt-Shin test#

btccny.kpss<-ur.kpss(btccny, type = c("mu"), lags = c("long"))
summary(btccny.kpss)
btckrw.kpss<-ur.kpss(btckrw, type = c("mu"), lags = c("long"))
summary(btckrw.kpss)
btcjpy.kpss<-ur.kpss(btcjpy, type = c("mu"), lags = c("long"))
summary(btcjpy.kpss)

#Zivot-Andrews test#

btccny.za<-ur.za(btccny, model = c("intercept"), lag=1)
summary(btccny.za)
plot(btccny.za)
Dates[1186]
btckrw.za<-ur.za(btckrw, model = c("intercept"), lag=1)
summary(btckrw.za)
plot(btckrw.za)
btcjpy.za<-ur.za(btcjpy, model = c("intercept"), lag=1)
summary(btcjpy.za)
plot(btcjpy.za)

###Retuns series###

#Augmented Dickey-Fuller test#
```

```
rendbtccny.df<-ur.df(rendbtccny, type = c("drift"), selectlags = c("BIC"))
summary(rendbtccny.df)
rendbtckrw.df<-ur.df(rendbtckrw, type = c("drift"), selectlags = c("BIC"))
summary(rendbtckrw.df)
rendbtcjpy.df<-ur.df(rendbtcjpy, type = c("drift"), selectlags = c("BIC"))
summary(rendbtcjpy.df)

#Phillips-Perron test#

rendbtccny.pp<-ur.pp(rendbtccny, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(rendbtccny.pp)
rendbtckrw.pp<-ur.pp(rendbtckrw, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(rendbtckrw.pp)
rendbtcjpy.pp<-ur.pp(rendbtcjpy, type = c("Z-tau"), model = c("constant"), lags = c("long"))
summary(rendbtcjpy.pp)

#Kwiatkowski-Phillips-Schmidt-Shin test#

rendbtccny.kpss<-ur.kpss(rendbtccny, type = c("mu"), lags = c("long"))
summary(rendbtccny.kpss)
rendbtckrw.kpss<-ur.kpss(rendbtckrw, type = c("mu"), lags = c("long"))
summary(rendbtckrw.kpss)
rendbtcjpy.kpss<-ur.kpss(rendbtcjpy, type = c("mu"), lags = c("long"))
summary(rendbtcjpy.kpss)

#Zivot-Andrews test#

rendbtccny.za<-ur.za(rendbtccny, model = c("intercept"), lag=1)
summary(rendbtccny.za)
plot(rendbtccny.za)
diffDates[1094]
```

```
rendbtckrw.za<-ur.za(rendbtckrw, model = c("intercept"), lag=1)
summary(rendbtckrw.za)
plot(rendbtckrw.za)
rendbtcjpy.za<-ur.za(rendbtcjpy, model = c("intercept"), lag=1)
summary(rendbtcjpy.za)
plot(rendbtcjpy.za)

###########Estimating VAR model###########

#Optimal lag selection#

data_var<-cbind(rendbtccny, rendbtckrw, rendbtcjpy)
VARselect(data_var,lag.max=20)
VAR_1<-VAR(data_var, p=1)

#Portmanteau Q and Lagrange Multiplier tests#

archtest<-arch.test(VAR_1)
plot(archtest)
predict_VAR<-predict(VAR_1, n.ahead=50)
plot(predict_VAR)

#Multivariate Portmanteau adjusted test#

mptad<-serial.test(VAR_1, lags.pt = 14, type = "PT.adjusted")

#Granger causality test - Robust HC variance-covariance matrix#

causality(VAR_1, cause = "rendbtccny", vcov.=vcovHC(VAR_1))
causality(VAR_1, cause = "rendbtckrw", vcov.=vcovHC(VAR_1))
```

```
causality(VAR_1, cause = "rendbtcjpy", vcov.=vcovHC(VAR_1))

#Impulse response function (IRF)#

irf(VAR_1, n.ahead=10) #Change order variables inside vector to check impacts
plot(irf(VAR_1, n.ahead=10))
VAR_summary<-summary(VAR_1)
VAR_summary$covres
VAR_summary$corres
t(chol(VAR_summary$covres))

###########Cointegration###########

###Johansen test###

#Maximum Eigenvalue and Trace tests#

data_coint<-cbind(btccny, btckrw, btcjpy)
max_eigen_test<-ca.jo(data_coint, type="eigen", ecdet="const", K=2)
summary(max_eigen_test)
trace_test<-ca.jo(data_coint, type="trace", ecdet="const", K=2)
summary(trace_test)

#VECM estimation#

VECM<-cajorls(max_eigen_test, r =1)
summary(VECM$rlm)
summary(vecm$rlm)
VECM$beta
```

```
###ARDL test###

data_coint_ardl<-cbind(btccny, btckrw, btcjpy)

#Model - order selection#

models_ardl<- auto_ardl(btccny ~ btckrw + btcjpy, data = data_coint_ardl, max_order = 5)
models_ardl$top_orders
ardl_model<- models_ardl$best_model
summary(ardl_model)

#UEM-RECM estimation#

uecm_ardl_model<- uecm(ardl_model)
summary(uecm_ardl_model)
recm_ardl_model<- recm(uecm_ardl_model, case = 3) #Case 3: unrestricted intercepts and no trend.
summary(recm_ardl_model)

#RECM coefficients#

tail(recm_ardl_model$coefficients, 1)
recm_ardl_model$coefficients[2]

#Ramsey RESET test#

resettest(ardl_model)

#Durbin-Watson test#

dwtest(recm_ardl_model, alternative = "two.sided")
```

```
#Breusch-Pagan test#

bptest(recm_ardl_model)

#Bounds F test#

bounds_f_test(uecm_ardl_model, case = 3) #Case 3: unrestricted intercepts and no trend.
t_bounds <- bounds_t_test(uecm_ardl_model, case = 3, alpha = 0.05)
t_bounds$tab

#Multipliers (sensibility) and Cointegration equation#

multipliers(ardl_model)
ce_ardl_model <- coint_eq(ardl_model, case = 3) #Case 3: unrestricted intercepts and no trend.

#Plot projected vs observed CE#

projvsobs <- cbind.zoo(btccny = data_coint_ardl[,"btccny"], ce_ardl_model)
projvsobs_xts <- xts(projvsobs)
plot(projvsobs_xts, legend.loc = "topleft")
```